

DOCUMENT RESUME

ED 250 390

TM 840 741

AUTHOR Brown, R. L.
TITLE The Efficacy of Multidimensional Line-Printer Graphics for Cluster Recovery.
PUB DATE 24 Aug 84
NOTE 22p.; Paper presented at the Annual Meeting of the American Psychological Association (92nd, Toronto, Ontario, August 24-28, 1984).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cluster Grouping; *Computer Graphics; Computer Simulation; Data Analysis; Data Processing; Distance; Higher Education; *Input Output Devices; *Multivariate Analysis; *Validity
IDENTIFIERS Computer Printers; *Data Interpretation

ABSTRACT

The plotting of multivariate data using computer line-printers has become a popular means of quickly representing multidimensional data. While many plotting programs are available, there is a paucity of research regarding the validity and reliability of interpretations made by viewing such graphics. This study explores the validity of four multidimensional line-printer graphics by measuring the efficacy of cluster recovery. Simulated datasets were created with known cluster parameters and plotted using four graphic methods (facial plots, stars, boxes, and linear profiles). Subjects attempted to recover dichotomous clusters from viewing the plots. Their performance was evaluated over levels of cluster difficulty (as indicated by different Euclidean distances), and levels of dimensionality. Results for one experiment (univariate study of complete link cluster method using dichotomous cluster recovery with no reliability assessment) indicated that the graphic techniques vary in the efficacy of cluster recovery, contingent upon level of cluster distance, and number of dimensions plotted. Results of a second experiment (multivariate study using dichotomous recovery with no reliability assessment) primarily indicated a graphic type by distance interaction. (Author)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED250390

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as
received from the person or organization
originating it.
Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

R. L. Brown

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

THE EFFICACY OF MULTIDIMENSIONAL
LINE-PRINTER GRAPHICS FOR CLUSTER RECOVERY

Dr. R. L. Brown
Center for Research on Pre-Traumatic Behavior
University of Wisconsin at Whitewater
Whitewater, Wisconsin 53190

Paper presented at the meeting of the American Psychological
Association, Toronto, Canada, August 24, 1984.

TM 840 741

ABSTRACT

THE EFFICACY OF MULTIDIMENSIONAL LINE-PRINTER GRAPHICS FOR CLUSTER RECOVERY

THE PLOTTING OF MULTIVARIATE DATA USING COMPUTER LINE-PRINTERS HAS BECOME A POPULAR MEANS OF QUICKLY REPRESENTING MULTIDIMENSIONAL DATA. WHILE MANY PLOTTING PROGRAMS ARE AVAILABLE, THERE IS A PAUCITY OF RESEARCH REGARDING THE VALIDITY AND RELIABILITY OF INTERPRETATIONS MADE BY VIEWING SUCH GRAPHICS. THIS STUDY EXPLORES THE VALIDITY OF FOUR MULTIDIMENSIONAL LINE-PRINTER GRAPHICS BY MEASURING THE EFFICACY OF CLUSTER RECOVERY.

SIMULATED DATASETS WERE CREATED WITH KNOWN CLUSTER PARAMETERS AND PLOTTED USING FOUR GRAPHIC METHODS (FACIAL PLOTS, STARS, BOXES, AND LINEAR PROFILES). SUBJECTS ATTEMPTED TO RECOVER DICHOTOMOUS CLUSTERS FROM VIEWING THE PLOTS. THEIR PERFORMANCE WAS EVALUATED OVER LEVELS OF CLUSTER DIFFICULTY (AS INDICATED BY DIFFERENT EUCLIDEAN DISTANCES), AND LEVELS OF DIMENSIONALITY.

RESULTS FOR THE FIRST OF FOUR EXPERIMENTS (UNIVARIATE STUDY OF COMPLETE LINK CLUSTER METHOD USING DICHOTOMOUS CLUSTER RECOVERY WITH NO RELIABILITY ASSESSMENT) INDICATE THAT THE GRAPHIC TECHNIQUES VARY IN THE EFFICACY OF CLUSTER RECOVERY, CONTINGENT UPON LEVEL OF CLUSTER DISTANCE, AND NUMBER OF DIMENSIONS PLOTTED.

RESULTS OF THE SECOND EXPERIMENT (MULTIVARIATE STUDY USING DICHOTOMOUS RECOVERY WITH NO RELIABILITY ASSESSMENT) PRIMARILY INDICATE A GRAPHIC TYPE BY DISTANCE INTERACTION.

THE EFFICACY OF MULTIDIMENSIONAL LINE-PRINTER GRAPHICS FOR CLUSTER RECOVERY

THE CURRENT PRACTICE OF DATA ANALYSIS CALLS FOR MORE AND MORE OF WHAT JOHN TUKEY HAS TERMED "EXPLORATORY DATA ANALYSIS" (1977). THIS POSITION CONCENTRATES ON "LOOKING" AT THE DATA TO DEVELOP INSIGHTS BY USING EASILY INTERPRETABLE GRAPHICS (TUKEY, 1977; TUFTE, 1983). EXPLORATORY DATA ANALYSIS HAS SUBSEQUENTLY STIMULATED A MYRIAD OF GRAPHIC ROUTINES, PROVIDING ANALYSTS WITH THE OPPORTUNITY TO VISUALLY EXPLORE BIVARIATE AS WELL AS MULTIVARIATE DATASETS (BARNETT, 1981; WAINER AND THISSEN, 1981; CHAMBERS, ET AL, 1983; TUFTE, 1983).

WHILE EXPLORATORY GRAPHICS ARE ON THE INCREASE, CONCERN HAS RECENTLY BEEN VOICED FOR THE NEED TO EVALUATE BOTH THE ACCURACY OF DATA REPRESENTATION (EGETH, 1981; BROWN, 1983A; 1983B; HAMNER AND TURNER, 1983; FRENI-TITULAER AND LOUV, 1984), AS WELL AS THE EFFECT OF IRRELEVANCY OF GRAPHIC COMPONENTS ON INTERPRETATION (NAVEH-BENJAMIN AND PACHELLA, 1982; TUFTE, 1983).

WHILE MANY OF THESE TECHNIQUES EXIST, THERE IS LITTLE DOCUMENTED EVIDENCE OF HOW "GOOD" THESE GRAPHICS ARE AT WHAT THEY PROFESS TO ASSESS (EGETH, 1981). A FEW STUDIES HAVE EVALUATED THE VALIDITY OF SPECIFIC MULTIVARIATE GRAPHICS (CHERNOFF AND RISVI, 1975; NAVEH-BENJAMIN AND PACHELLA, 1982, BROWN, 1983C:), WITH SOME STUDIES COMPARING EFFICACY RATES ACROSS VARIOUS MULTIVARIATE GRAPHIC TECHNIQUES (MEZZICH AND WORTHINGTON, 1978; HAMNER AND TURNER, 1983; FRENI-TITULAER AND LOUV, 1984).

WHILE MULTIVARIATE GRAPHIC TECHNIQUES ARE READILY AVAILABLE, BOTH FOR X-Y PLOTTERS, AND LINE-PRINTERS, THERE IS A PAUCITY OF RESEARCH AVAILABLE ON THE EFFICACY OF MAKING DATA INTERPRETATIONS USING LINE PRINTER PRODUCED GRAPHICS.

THE PURPOSE OF THIS RESEARCH IS TO ESTABLISH CLUSTER RECOVERY EFFICACY RATES ON FOUR POPULAR MULTIDIMENSIONAL LINE-PRINTER GRAPHIC TECHNIQUES USED IN EXPLORATORY DATA ANALYSIS. THE GRAPHIC TECHNIQUES BEING EXPLORED ARE: (1) CARTOON-LIKE FACIAL PLOTS (CHERNOFF, 1973), (2) THREE DIMENSIONAL BOX PLOTS (HARTIGAN, 1975), (3) STAR PLOTS (FRIEDMAN, FARRELL, GOLDWYN, MILLER AND SIEGEL, 1972), AND (4) LINEAR PROFILES (MEZZICH AND WORTHINGTON, 1978), (SEE FIGURE 1A,B,C,D).

RESEARCH QUESTIONS

A. WHAT EFFECT WILL CLUSTER DISTANCE (MEASURED IN EUCLIDEAN DISTANCE) HAVE ON DICHOTOMOUS CLUSTER RECOVERY USING THE ABOVE MENTIONED GRAPHICS?

B. WHAT EFFECT WILL DIMENSIONALITY HAVE ON DICHOTOMOUS CLUSTER RECOVERY?

C. WHAT ARE THE INTERACTIONS BETWEEN TYPES OF GRAPHIC USED FOR DICHOTOMOUS CLUSTER RECOVERY, AND DISTANCE OF PRE-ESTABLISHED DATA MATRICES AND DIMENSIONALITY?

METHODS

CLUSTER SIMULATIONS. A COMPUTER SIMULATION WAS WRITTEN THAT CONSTRUCTS TWO EXPERIMENTAL MATRICES ($\tilde{X}_{1n,k}$ AND $\tilde{X}_{2n,k}$) OF ORDER (10 X 4), WITH EACH MATRIX PARTITIONED INTO TWO SUBMATRICES OF ORDER N_i X K, WHERE ($i=1,2$), AND $3 < N_1 < 6$, $N_2 = 10 - N_1$, AND $K=6$. SUBMATRICES WERE COMPOSED OF SIX COLUMN VECTORS EACH ($\tilde{X}_1 - \tilde{X}_6$) AND SAMPLED FROM TWO POPULATIONS OF NORMALLY DISTRIBUTED PSEUDO-RANDOM VARIATES, WITH μ 'S = 50.0 TO 54.0 AND $\sigma = 5.0$. THE SIMULATION THUS FORCED TWO CLUSTERS PER EXPERIMENTAL $\tilde{X}$ MATRIX (SUBMATRICES $\tilde{X}_{1s}$ AND $\tilde{X}_{2s}$). DISTANCE (EUCLIDEAN METRIC δ) BETWEEN THE TWO SUBMATRICES (CLUSTERS) WAS VARIED TO PRODUCE TWO DISTANCE CONDITIONS (ONE SMALL AND ONE LARGE). DIMENSIONALITY OF THE DATASETS WERE GOVERNED BY SIMPLY DUPLICATING THE 10 X 4 MATRIX (SMALL DIMENSION) AND RANDOMLY COMBINING THE TWO DUPLICATES INTO AN 10 X 8 MATRIX (LARGE DIMENSION). THE FLOWCHART IN FIGURE 2 DEMONSTRATES THE CLUSTER SIMULATION PROCESS FOR ESTABLISHING SMALL AND LARGE DISTANCE CONDITIONS.

PROCEDURES.

EXPERIMENT NUMBER 1: EIGHTY STUDENTS WERE RECRUITED FROM THE UNIVERSITY OF WISCONSIN CAMPUS. EACH STUDENT WAS GIVEN A BRIEF TRAINING SESSION REGARDING THE USE OF MULTIDIMENSIONAL GRAPHICS. STUDENTS WERE THEN ASKED TO SORT A PACKET OF THREE DIFFERENT GRAPHICS (FACIAL PLOTS, LINEAR PROFILES AND 3D BOX PLOTS) INTO ONE OF TWO GROUPS (DICHOTOMOUS CLUSTER RECOVERY). THE STUDENTS WERE ALSO REQUESTED TO CLUSTER A SET OF RAW DATA.

EXPERIMENT NUMBER 2: FOURTEEN STUDENTS WERE RECRUITED FROM THE UNIVERSITY OF WISCONSIN CAMPUS TO PARTICIPATE IN THIS EXPERIMENT. STUDENTS WERE TRAINED IN THE USE OF THE VARIOUS MULTIDIMENSIONAL GRAPHICS. STUDENTS WERE THEN REQUIRED TO SORT GRAPHICS SIMILAR TO THE PROCEDURE EXPLAINED IN THE FIRST EXPERIMENT (WITH THE ADDITION OF THE STAR GRAPHIC).

DATA COLLECTION:

THE CLUSTER RECOVERY DATA WAS COLLECTED FROM THE SORTED PACKETS OF STIMULUS MATERIALS (GRAPHICS) AND CODED VIA A SPECIAL COMPUTER CODING PROGRAM. EFFICACY RATES WERE THEN ESTABLISHED BY COMPARING SUBJECT TO COMPUTER CLUSTER RECOVERY FOR DICHOTOMOUS CLUSTERS USING FIVE DIFFERENT CLUSTER RECOVERY METHODS (E.G., BETWEEN AVERAGE, WITHIN AVERAGE, CENTROID, SINGLE LINK AND COMPLETE LINK). A MEASURE OF ASSOCIATION BETWEEN THE STUDENT'S CLUSTER RECOVERY AND THE COMPUTER'S WAS CALCULATED ($\phi^2 = X^2/N$).

STATISTICAL ANALYSES

EXPERIMENT NUMBER ONE USED AN ANOVR 2-2-4 SPLIT PLOT FACTORIAL DESIGN, WITH TWO BETWEEN FACTORS OF DIMENSIONALITY, TWO BETWEEN FACTORS OF DISTANCE AND FOUR WITHIN FACTORS OF GRAPHIC TYPE. EXPERIMENT NUMBER TWO USED A MANOVR 2-2-5 WITHIN DESIGN, WITH TWO WITHIN FACTORS OF DIMENSIONALITY, TWO WITHIN FACTORS OF DISTANCE, AND FIVE WITHIN FACTORS OF GRAPHIC TYPE.

RESULTS

THE FINDINGS FROM EXPERIMENT 1 INDICATE THAT MULTIVARIATE LINE-PRINTER GRAPHICS VARY IN EFFICACY OF DICHOTOMOUS CLUSTER RECOVERY FOR THE COMPLETE LINK METHOD OF CLUSTERING, CONTINGENT UPON CLUSTER DISTANCE AND NUMBER OF DIMENSIONS (SEE TABLE 1 AND FIGURE 3).

EXPERIMENT 2 ASSESSED RECOVERY EFFICACY OVER FIVE DEPENDENT CLUSTER RECOVERY METHODS: (1) BETWEEN AVERAGE, (2) WITHIN AVERAGE, (3) CENTROID, (4) SINGLE LINK, AND (5) COMPLETE LINK. A MANOVR WAS CONDUCTED (SEE TABLE 2). A SIGNIFICANT GRAPHIC TYPE BY DISTANCE INTERACTION WAS DETECTED (SEE MULTIPLE UNIVARIATE INTERACTION PLOTS, FIGURE 4A,B,C,D). A POST-COMPARISON MODIFIED DISCRIMINANT ANALYSIS (MDA) WAS CONDUCTED ON THE MANOVR (TATSUOKA, 1971). FIGURE 5 IS A PLOT OF THE TWO SIGNIFICANT DISCRIMINANT FUNCTIONS ON THE GRAPHIC TYPE BY DISTANCE INTERACTION. IT IS EASILY SEEN THAT THE FACIAL PLOT (IN BOTH DISTANCE CONDITIONS) DIFFER FROM THE OTHER GRAPHICS, WITH THE EXCEPTION OF THE SMALL DISTANCE LINEAR PLOTS.

CONCLUSIONS

DISTANCE BETWEEN CLUSTER STRUCTURES INTERACT WITH THE TYPE OF GRAPHIC METHOD USED OVER A VARIETY OF CLUSTER INTERPRETATIONAL METHODS. HARTIGAN'S (1975) BOX PLOTS AND THE STAR PLOTS PROVIDED THE LEAST EFFICIENCY IN DICHOTOMOUS CLUSTER RECOVERY ACROSS ALL DEPENDENT CONDITIONS, REGARDLESS OF DISTANCE EFFECT. THE POPULAR FACIAL PLOT PROVIDED A SOMEWHAT CONFUSING PICTURE. IF CLUSTER RECOVERY WAS BASED ON THE NEARNESS OF THE CLUSTER STRUCTURES (E.G., SINGLE LINK, CENTROID, OR BETWEEN AVERAGE METHODS) THE FACIAL PLOT INTERPRETATION WAS STRONGLY INFLUENCED BY THE DISTANCE BETWEEN SUCH STRUCTURES, WHILE

CLUSTER RECOVERY BASED ON TIGHTNESS OF CLUSTERS (E.G., WITHIN AVERAGE METHOD), AND FURTHEST STRUCTURES (COMPLETE LINK METHOD) PROVIDED BEST CLUSTER RECOVERY USING THE FACIAL PLOTS, REGARDLESS OF DISTANCE.

ALSO WORTHY OF NOTING WAS THAT SIMPLE LINEAR PROFILES AND INTERPRETATION FROM THE RAW DATA MATRICES GENERALLY PROVIDED BETTER RECOVERY THAN EITHER BOX OR STAR PLOTS, BUT WERE INFLUENCED BY DISTANCE.

THIS STUDY WAS LIMITED TO DICHOTOMOUS CLUSTER RECOVERY WHICH MAY HAVE MADE RECOVERY EASIER WITH SOME GRAPHICS THAN WITH OTHERS. TWO ONGOING EXPERIMENTS WILL LOOK AT DYNAMIC (>2) CLUSTER RECOVERY, WITH AN IMPORTANT RELIABILITY FACTOR FOR EACH GRAPHIC TECHNIQUE.

FIGURE 1A

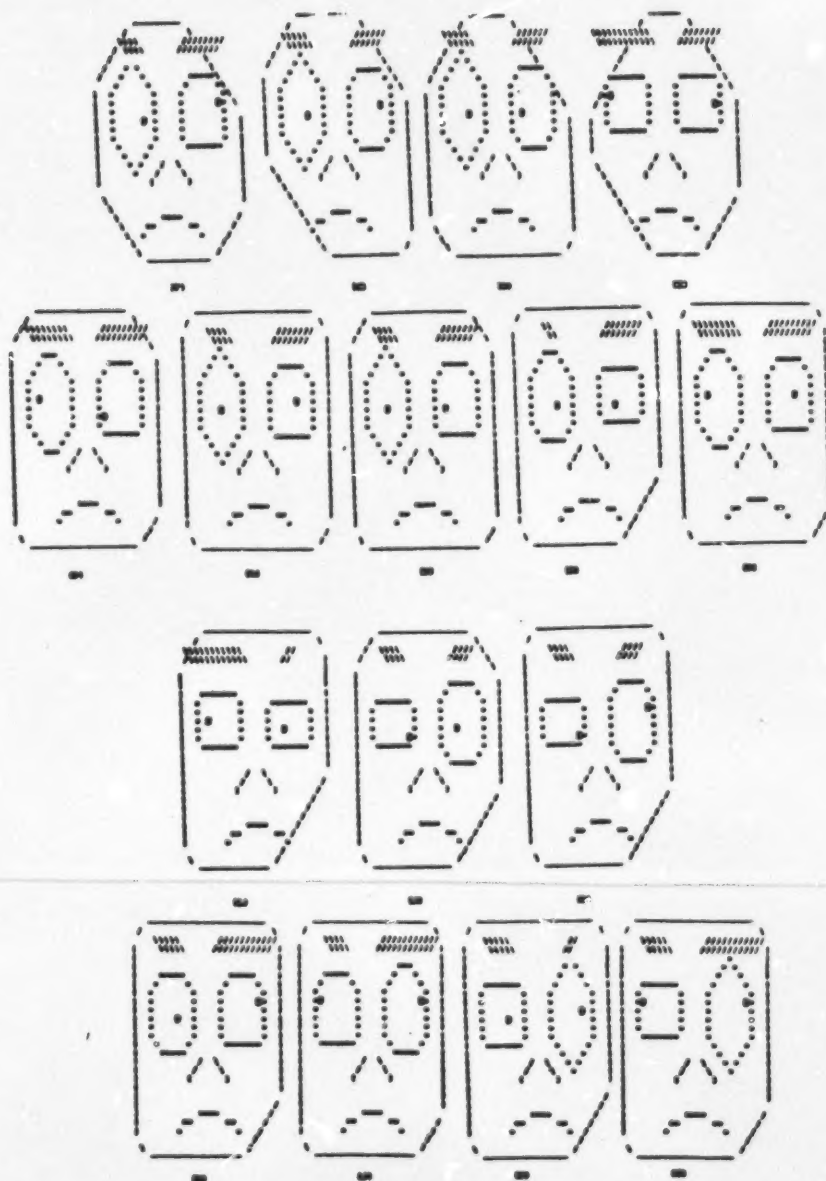


FIGURE 1B

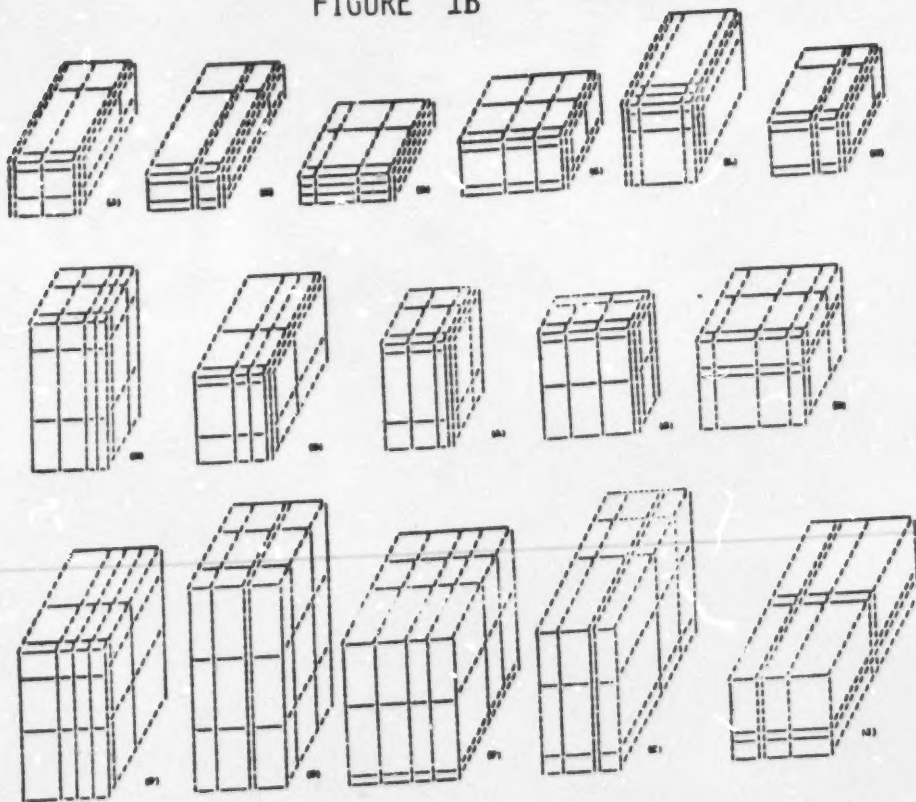


FIGURE 1C

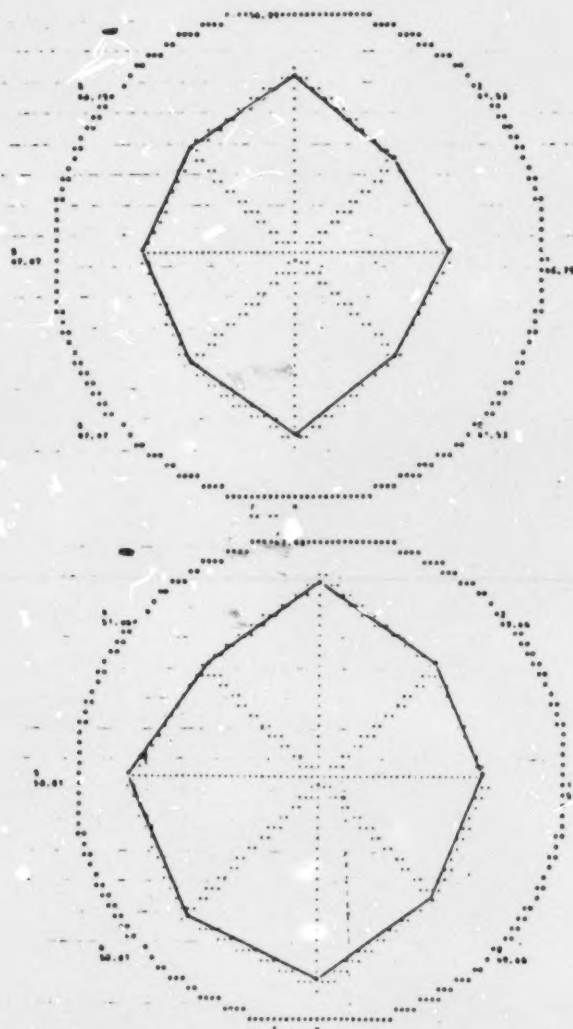


FIGURE 1D

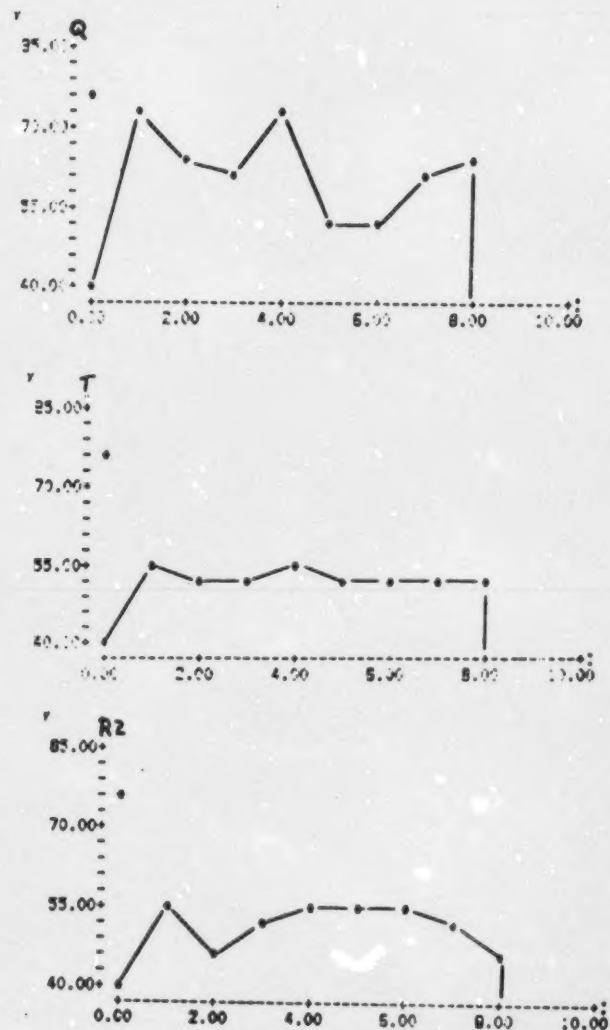
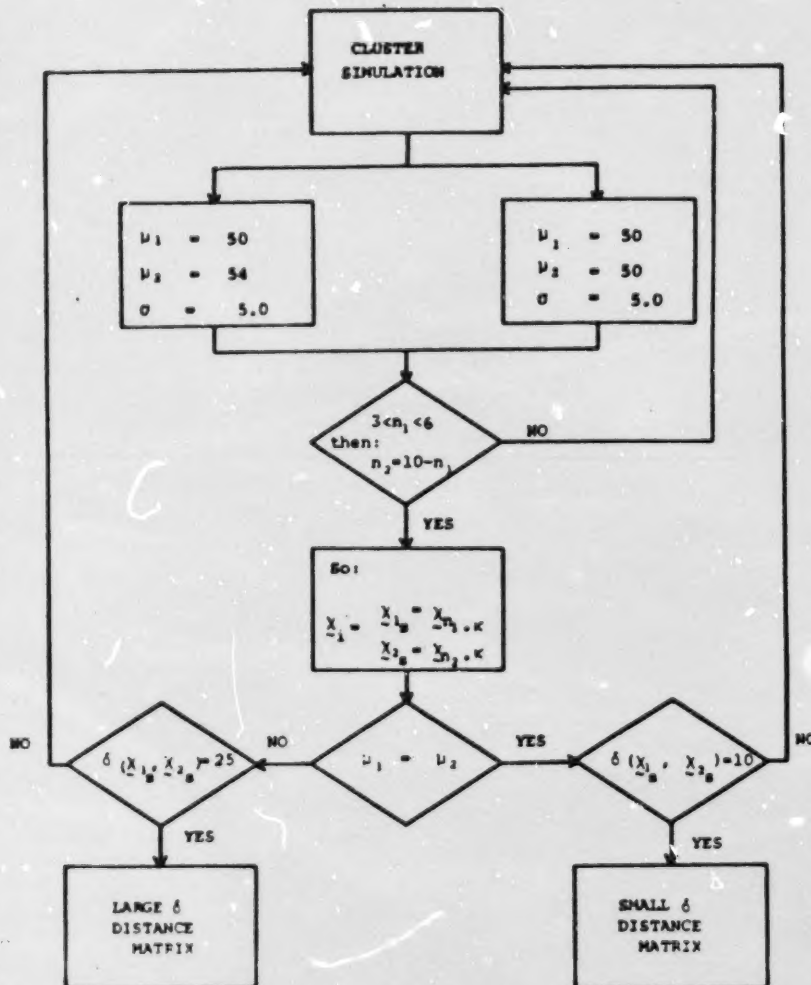


FIGURE 2



BEST COPY AVAILABLE

TABLE 1
UNIVARIATE ANALYSIS FOR EXPERIMENT NUMBER ONE
ANALYSIS OF VARIANCE SUMMARY TABLE

SOURCE	SUMS OF SQUARES	MEAN SQUARES	DF	F-RATIO
BETWEEN SUBJECTS				
DIMENSIONALITY (A)	.02048	.02048	1	0.540
DISTANCE (B)	.30135	.30135	1	7.952*
A X B INTERACTION	.00112	.00112	1	0.030
ERROR	2.87899	.03789	76	
WITHIN SUBJECTS				
GRAPHIC TYPES (C)	2.30493	.76831	3	23.193*
A X C INTERACTION	.94603	.31534	3	9.520*
B X C INTERACTION	.33299	.11099	3	3.351
A X B X C INTERACTION	2.08110	.69370	3	20.941*
ERROR	7.55275	.03313	228	

(* P < .05)

GRAPHIC BY DIMENSION BY DISTANCE INTERACTION COMPLETE LINK METHOD

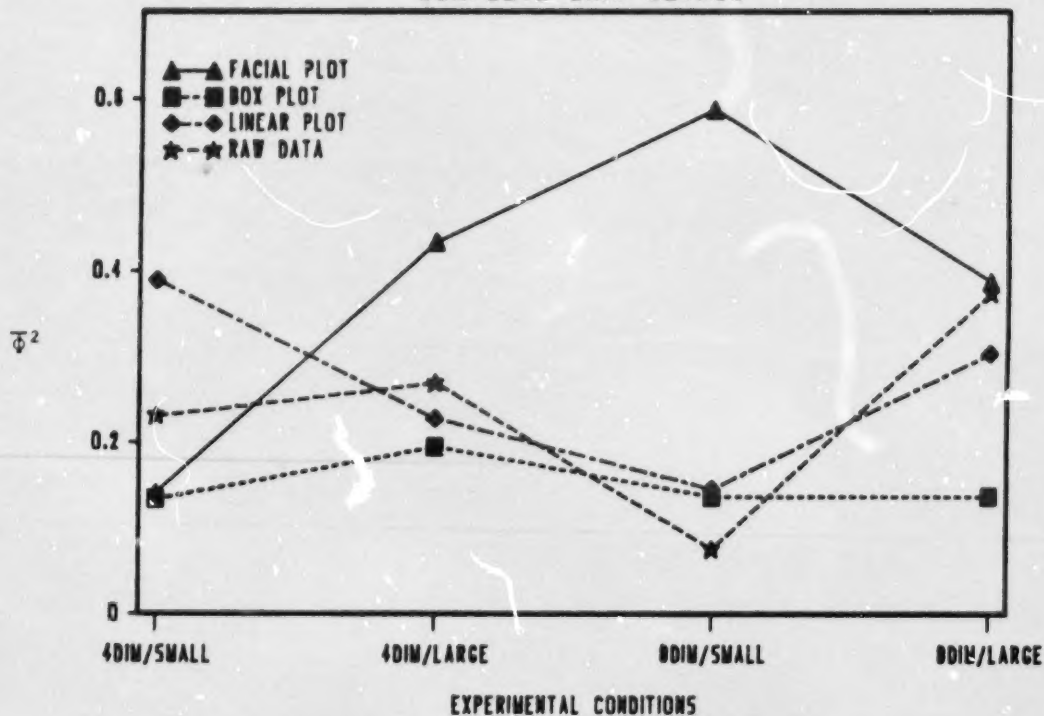


FIGURE 3

TABLE 2
MULTIVARIATE ANALYSIS FOR EXPERIMENT NUMBER TWO
MANOVR SUMMARY TABLE

SOURCE OF VARIANCE	DF	WILKS	APPROX F	PROB
MAIN EFFECTS				
DISTANCE (A)	5,9	.72417	< 1.0	.646
DIMENSIONALITY (B)	5,9	.92974	< 1.0	.980
GRAPHIC TYPE (C)	20,160	.61443	1.266	.209
INTERACTIONS				
(A) X (B)	5,9	.72725	< 1.0	.653
(A) X (C)	20,160	.44894	2.186	.004**
(B) X (C)	20,160	.73651	< 1.0	.742
(A) X (B) X (C)	20,160	.55877	1.535	.076

(** = $P < .05$)

GRAPHIC TYPE BY DISTANCE INTERACTION D.V. 1 : BETWEEN AVERAGE METHOD

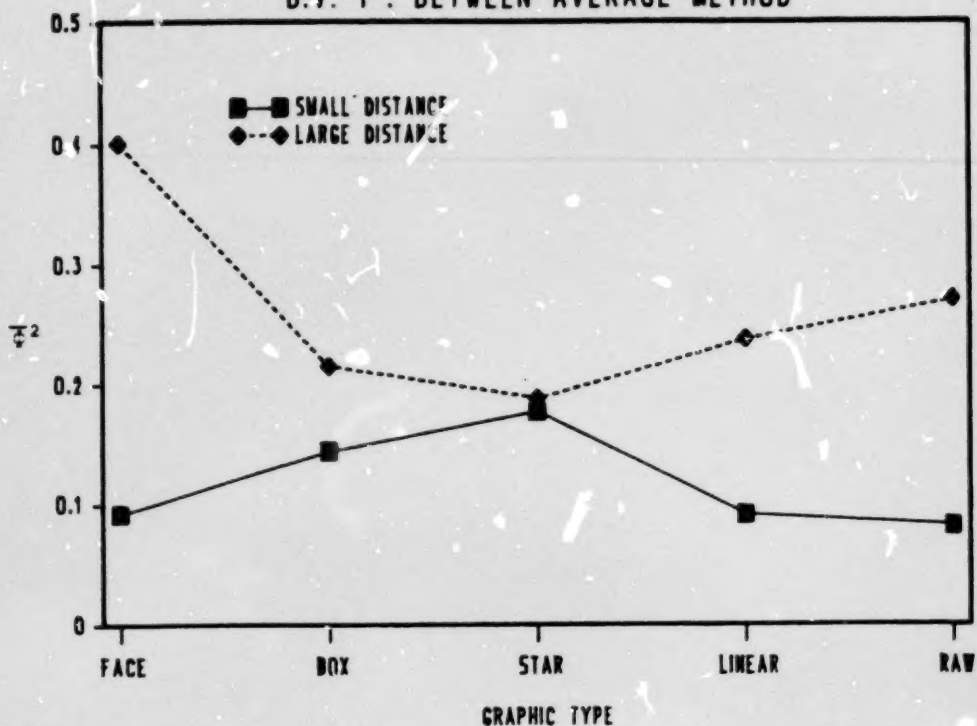
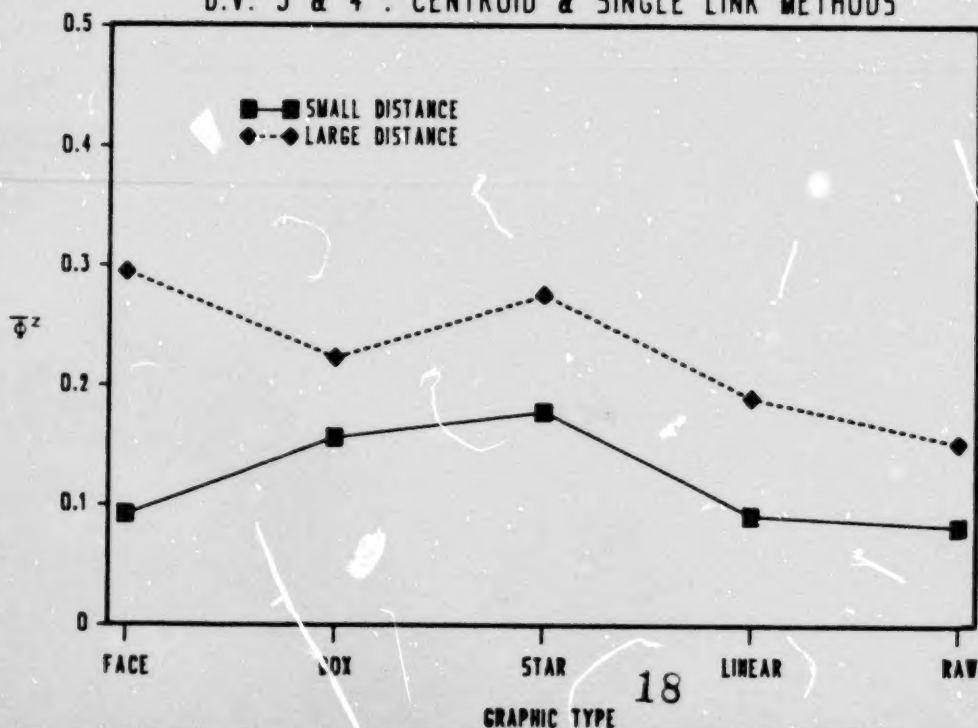
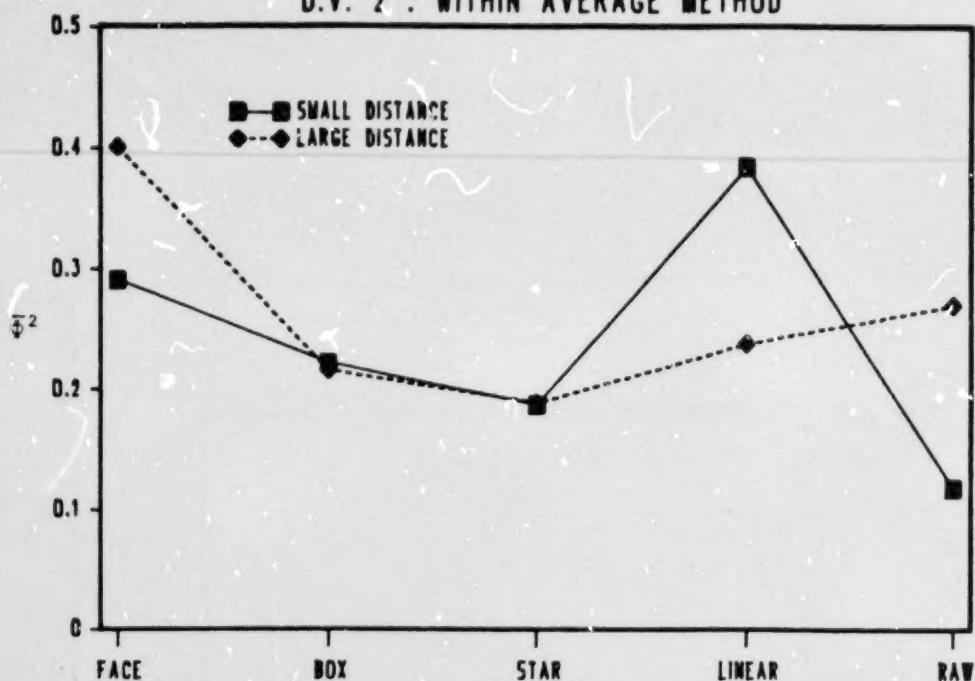


FIGURE 4A

GRAPHIC TYPE BY DISTANCE INTERACTION D.V. 3 & 4 : CENTROID & SINGLE LINK METHODS

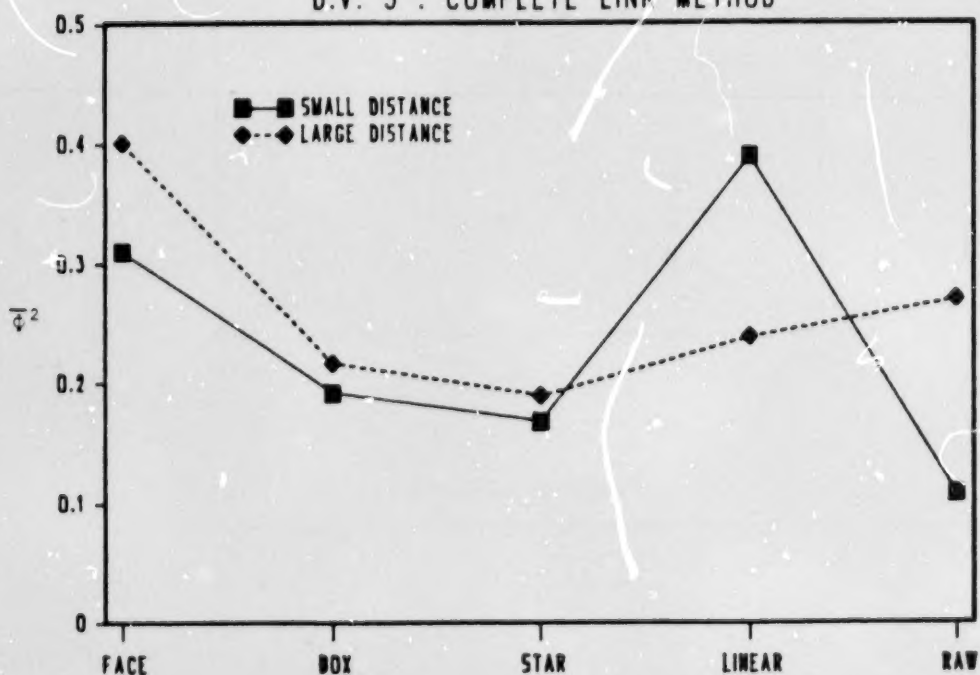


GRAPHIC TYPE BY DISTANCE INTERACTION D.V. 2 : WITHIN AVERAGE METHOD



GRAPHIC TYPE
FIGURE 4B

GRAPHIC TYPE BY DISTANCE INTERACTION D.V. 5 : COMPLETE LINK METHOD



GRAPHIC TYPE

MODIFIED DISCRIMINANT ANALYSIS (MDA) POST-COMPARISON MANOVR

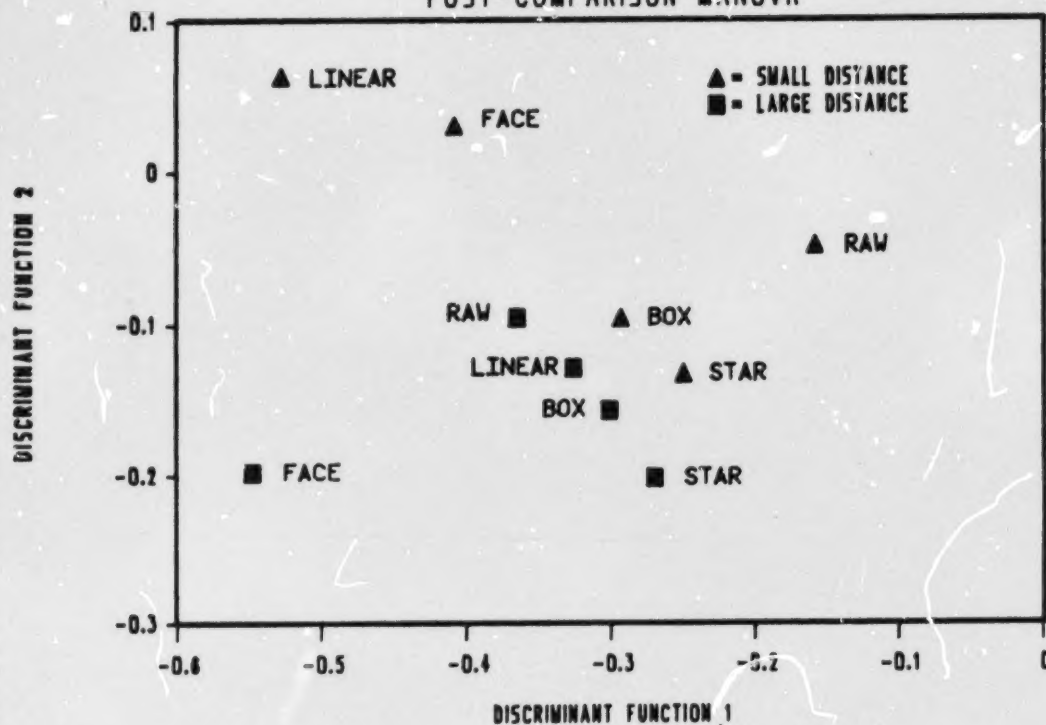


FIGURE 5

BEST COPY AVAILABLE

REFERENCES

Barnett, V. (Ed.). Interpreting multivariate data. New York: John Wiley and Sons, 1981.

Brown, R. L. Multidimensional line printer graphics for use in system safety. Proceedings of the sixth International System Safety Conference, 1983a, 1, 4.2.1-4.2.19.

Brown, R. L. Multidimensional graphic techniques for exploratory analysis of archival accident data. Accident Analysis and Prevention, 1983b, 15(4), 299-308.

Brown, R. L. The influence of the mouth on the clustering of facial plots. Paper presented at the meeting of the Ninety-first Annual Convention of the American Psychological Association, Anaheim, August 1983c.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. Graphical methods for data analysis. Belmont, California: Wadsworth International Group, 1983.

Chernoff, H. The use of faces to represent points in n-dimensional space graphically. Stanford University, California: Department of Statistics, December, 1971. (Technical Report No. 71).

Chernoff, H., and Rizvi, M. H. Effect on classification error of random permutations of features in representing multivariate data by faces. Journal of the American Statistical Association, 1975, 70, 542-554.

Egeth, H. Comment on tress and castles. Journal of the American Statistical Association 1981, 76(374), 269-270.

Freni-Titulaer, L. W. J., and Louv, W. C. Comparisons of some graphical methods for exploratory multivariate data analysis. American Statistician. 1984, 38, 184-188.

Friedman, H. P., Farrell, E. S., Goldwyn, R. M., Miller, M., and Siegel, J. H. A graphic way of describing changing multivariate patterns. Proceedings of the Sixth Interface Symposium on Computer Science and Statistics, 1972, 56-59.

Hamner, C. G., and Turner, D. W. A study comparing several graphical methods for representing multivariate data. Paper presented at the meeting of the American Statistical Association, Toronto, Canada, August, 1983.

Hartigan, J. A. Printer graphics for clustering. Journal of Statistics and Computer Simulation 1975, 4, 187-213.

Mezzich, J. E., and Worthington, D. R. L., A comparison of graphical representations of multidimensional psychiatric diagnostic data. In P. C. C. Wang (Ed.), Graphical representation of multivariate data. New York: Academic Press, 1978.

Naveh-Benjamin, M., and Pachella, R. G. The effect of complexity on interpreting "Chernoff" faces. Human Factors, 1982, 24(1), 11-18.

Tatsuoka, M. M. Multivariate analysis. New York: John Wiley and Sons, 1971.

Tufte, E. R. The visual display of quantitative information. Cheshire, Connecticut: Graphics Press, 1983.

Tukey, J. W. Exploratory data analysis. Reading, Massachusetts: Addison-Wesley, 1977.

Wainer, H., and Thissen, D. Graphical data analysis. Annual Review of Psychology, 1981, 32, 191-241.